Predictors of Earnings Risk with Machine Learning

Ethan Ballou* Scott Drewianka*

December 10, 2025

Abstract

This paper looks at the determinants of lifetime earnings risk under a Restricted Income Profile (RIP) model using traditional and machine learning methods such as lasso and SHAP values. The paper builds on the work of Drewianka and Oberg (2025) which uses a moment condition approach derive a parameter that captures permanent income risk. The paper finds that education and age are important in explaining lifetime earnings risk. The paper also finds that macroeconomic variables such as probability of recession and real GDP growth are important and along with state controls may further imply a role of government policy. Finally, the paper finds that occupation controls are important while industry controls do not appear to play a strong role.

Keywords: machine learning, restricted income profile, earnings instability, risk
JEL Codes: D8, J0, D3

*University of Wisconsin - Milwaukee

1. NEED TO ADD INFO ABOUT WHICH DUMMIES WERE INCLUDED IN STEPWISE REGRESSIONS SINCE THEY ARE SUPPRESSED ON THE TABLE
2. NEED TO ADD IN AND OCD Lasso, and of all of lasso for that matter

1 Introduction

2 Data and Model

3 Empirical Strategy

4 Results

5 Conclusion

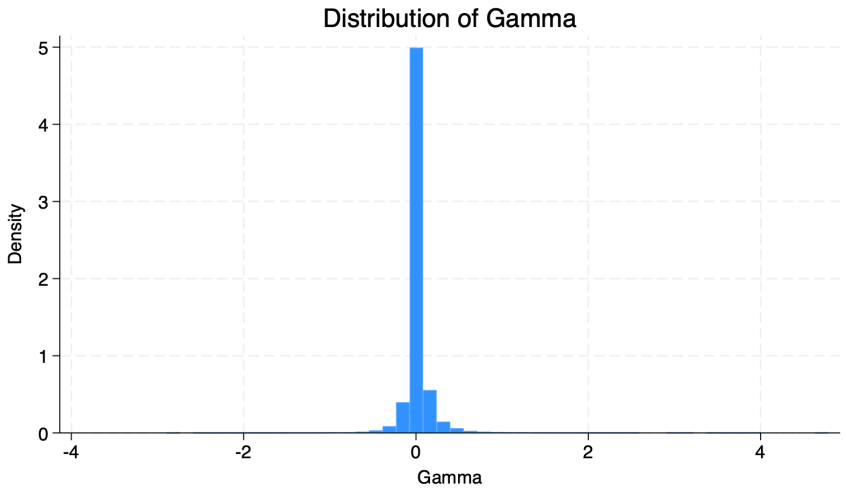

Figure 1: Distribution of Gamma

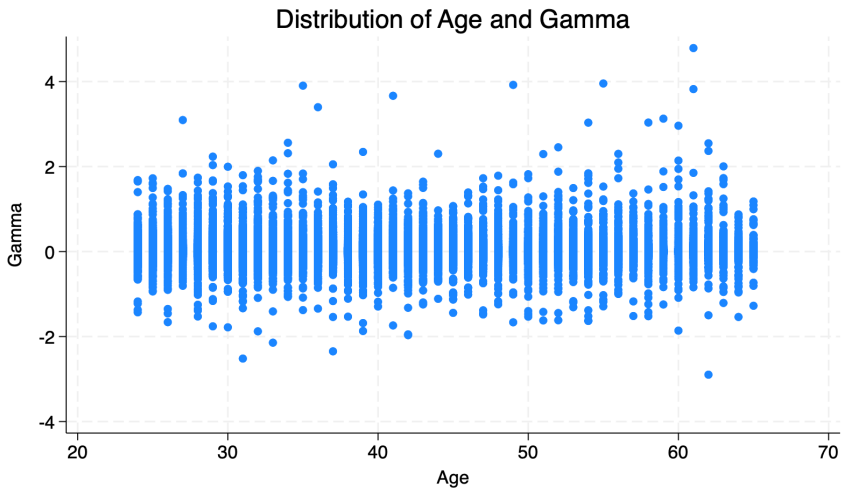

Figure 2: Scatterplot of Age vs. Gamma

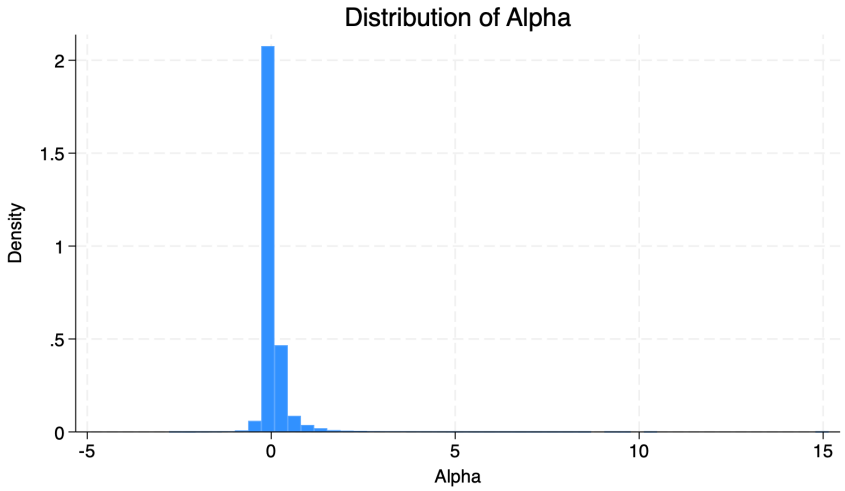

Figure 3: Distribution of Alpha

Table 1: Gamma Regressions: OLS Results

	(1) Gamma	(2) Gamma	(3) Gamma	(4) Gamma	(5) Gamma
Less than High School	-0.00950*	-0.00719	-0.00542	-0.00196	-0.00479
	(0.00407)	(0.00437)	(0.00485)	(0.00523)	(0.00519)
High School Graduate	-0.00632*	-0.00591*	-0.00506	-0.00287	-0.00399
	(0.00259)	(0.00269)	(0.00294)	(0.00334)	(0.00331)
Some College	-0.00278	-0.00277	-0.00186	-0.000549	-0.000961
	(0.00296)	(0.00304)	(0.00323)	(0.00350)	(0.00345)
Probability of Recession	0.0000411	-0.00207	-0.0962	-0.000140	-0.0953
	(0.0000842)	(247.5)	(0.0528)	(297.1)	(0.0550)
Real GDP growth rate	0.000553	-0.0130	-0.0137	-0.00832	-0.0161*
	(0.000719)	(19.13)	(0.00703)	(22.96)	(0.00722)
5-year moving average of AEP	0.0000266	0.000000171	-0.00000250	0.0000340	0.0000309
	(0.0000548)	(0.0000595)	(0.0000654)	(0.0000693)	(0.0000700)
Veteran	-0.00281	-0.00341	-0.00610	-0.00755	-0.00828
	(0.00436)	(0.00447)	(0.00649)	(0.00686)	(0.00677)
Out of Labor Force	-0.0101	-0.00862	-0.0227	-0.0216	-0.0220
	(0.00848)	(0.00853)	(0.0142)	(0.0146)	(0.0144)
Tenure	0.0000614	0.000140	0.000202	0.000151	0.000184
	(0.000151)	(0.000162)	(0.000164)	(0.000171)	(0.000168)
Age	0.00592	0.00559	0.00559	0.00884	0.00733
	(0.00491)	(0.00504)	(0.00558)	(0.00581)	(0.00567)
Age Squared	-0.000160	-0.000169	-0.000172	-0.000252	-0.000213
	(0.000116)	(0.000119)	(0.000130)	(0.000135)	(0.000132)
Age Cubed	0.00000140	0.00000150	0.00000155	0.00000219*	0.00000185
	(0.000000886)	(0.000000906)	(0.000000984)	(0.00000102)	(0.00000100)
State FE	No	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	Yes	Yes
Industry FE	No	No	Yes	No	Yes
R-squared	0.001	0.004	0.009	0.010	0.013
N	25920	25913	20954	20823	20746

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Gamma Regressions: Stepwise Selection Results

	(1) Gamma	(2) Gamma	(3) Gamma	(4) Gamma	(5) Gamma
Less than High School	-0.00768*	-0.00769*			
	(0.00363)	(0.00364)			
High School Graduate	-0.00489*	-0.00494*			
	(0.00212)	(0.00212)			
Age Cubed	0.000000364*	0.000000367*	0.000000449**	0.000000558***	0.000000476**
	(0.000000145)	(0.000000145)	(0.000000157)	(0.000000162)	(0.000000158)
Age Squared	-0.0000222*	-0.0000225*	-0.0000287**	-0.0000352**	-0.0000304**
	(0.00000985)	(0.00000985)	(0.0000108)	(0.0000111)	(0.0000109)
Constant	0.0304***	0.0306***	0.0292***	0.0349***	0.0302***
	(0.00619)	(0.00619)	(0.00726)	(0.00710)	(0.00732)
State FE					
Year FE					
Race FE					
Cohort FE					
Occupation FE					
Industry FE			✓		✓
State FE Available	No	Yes	Yes	Yes	Yes
Year FE Available	No	Yes	Yes	Yes	Yes
Race FE Available	No	Yes	Yes	Yes	Yes
Cohort FE Available	No	Yes	Yes	Yes	Yes
Occupation FE Available	No	No	No	Yes	Yes
Industry FE Available	No	No	Yes	No	Yes
R-squared	0.001	0.001	0.004	0.001	0.004
N	25920	25913	20954	20823	20746

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Gamma Regressions: Lasso Selection Results

	No Controls	No Occ/Ind	No Occ	No Ind	All Controls
EDU1	3	4	5	8	5
EDU2	2	2	3	4	4
EDU3	8	7	8	9	8
OLF	4	5	2	3	3
PrRecess	11				
currentage	10	6	7	7	6
currentagecube	1	8	6	6	7
currentagesq	12	9	10	10	9
ma5aep	7	10	9	5	5
rGDPgrow	6				
tenure	9	6	4	6	5
veteran	5	3	3	2	2
State FE	No	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	Yes	Yes
Industry FE	No	No	Yes	No	Yes

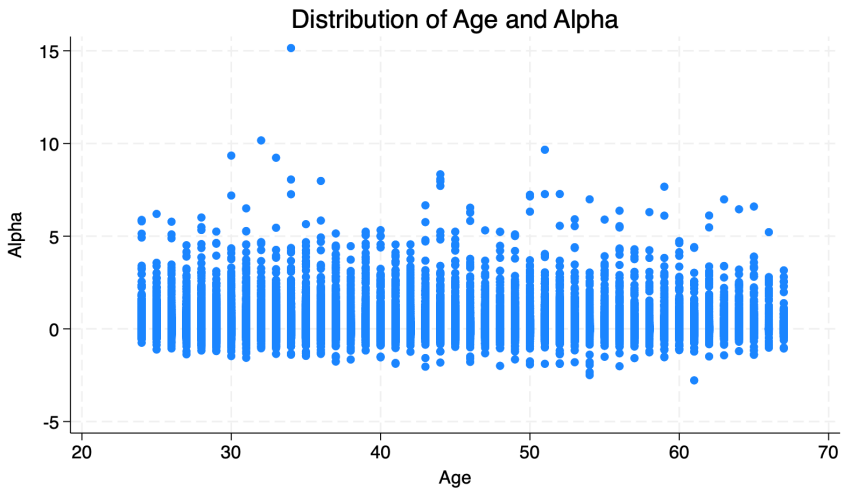

Figure 4: Scatterplot of Age vs. Alpha

Table 4: Alpha Regressions: OLS Results

	(1) Alpha	(2) Alpha	(3) Alpha	(4) Alpha	(5) Alpha
Less than High School	-0.0442***	-0.0342***	-0.0299**	-0.0285*	-0.0271*
	(0.00889)	(0.00954)	(0.0108)	(0.0114)	(0.0115)
High School Graduate	-0.0387***	-0.0364***	-0.0341***	-0.0301***	-0.0277***
	(0.00580)	(0.00600)	(0.00664)	(0.00736)	(0.00743)
Some College	-0.0234***	-0.0242***	-0.0184*	-0.0162*	-0.0146
	(0.00666)	(0.00681)	(0.00732)	(0.00774)	(0.00778)
Probability of Recession	-0.0000102	-0.0449	0.00151	0.255	0.245
	(0.000185)	(0.167)	(664.2)	(0.216)	(0.216)
Real GDP growth rate	-0.00108	-0.00306	0.0169	0.0130	0.0126
	(0.00158)	(0.0135)	(5.775)	(0.0175)	(0.0176)
5-year moving average of AEP	0.00142***	0.00153***	0.00129***	0.00148***	0.00144***
	(0.000119)	(0.000129)	(0.000144)	(0.000149)	(0.000154)
veteran	-0.0124	-0.0122	-0.00408	-0.000294	-0.00321
	(0.00920)	(0.00941)	(0.0146)	(0.0150)	(0.0152)
Out of Labor Force	0.122***	0.123***	0.0984***	0.0891**	0.0923**
	(0.0165)	(0.0165)	(0.0289)	(0.0291)	(0.0291)
Tenure	-0.000825*	-0.0000632	-0.00000483	-0.000294	-0.000244
	(0.000330)	(0.000355)	(0.000365)	(0.000373)	(0.000372)
Age	0.0158	0.0186	0.00916	0.0162	0.0113
	(0.00985)	(0.0101)	(0.0115)	(0.0116)	(0.0116)
Age Squared	-0.000433	-0.000533*	-0.000283	-0.000446	-0.000321
	(0.000229)	(0.000233)	(0.000263)	(0.000266)	(0.000266)
Age Cubed	0.00000394*	0.00000477**	0.00000278	0.00000403*	0.00000299
	(0.00000171)	(0.00000175)	(0.00000195)	(0.00000198)	(0.00000197)
State FE	No	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	Yes
Race FE	No	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	Yes
Occupation FE	No	No	No	Yes	Yes
Industry FE	No	No	Yes	No	Yes
R-squared	0.013	0.019	0.029	0.034	0.037
N	32567	32556	25592	25426	25329

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Alpha Regressions: Stepwise Selection Results

	(1) Alpha	(2) Alpha	(3) Alpha	(4) Alpha	(5) Alpha
Less than High School	-0.0456***	-0.0350***	-0.0308**	-0.0296**	
	(0.00877)	(0.00947)	(0.0107)	(0.0113)	
High School Graduate	-0.0395***	-0.0374***	-0.0346***	-0.0305***	-0.0156**
	(0.00578)	(0.00596)	(0.00657)	(0.00729)	(0.00538)
Some College	-0.0237***	-0.0255***	-0.0185*	-0.0158*	
	(0.00665)	(0.00677)	(0.00725)	(0.00769)	
Age Cubed	0.00000131***	0.00000176***	0.00000119***	0.00000129***	0.00000112**
	(0.000000303)	(0.000000329)	(0.000000338)	(0.000000344)	(0.000000343)
Age Squared	-0.0000734***	-0.000119***	-0.0000686**	-0.0000752**	-0.0000653**
	(0.0000212)	(0.0000248)	(0.0000239)	(0.0000243)	(0.0000242)
5-year moving average of AEP	0.00146***	0.00152***	0.00127***	0.00149***	0.00137***
	(0.000115)	(0.000121)	(0.000139)	(0.000144)	(0.000146)
Tenure	-0.000824*				
	(0.000329)				
Out of Labor Force	0.124***	0.126***	0.101***	0.0942**	0.0973***
	(0.0164)	(0.0164)	(0.0287)	(0.0289)	(0.0289)
Constant	0.0774***	0.154**	0.105	0.102	0.0986
	(0.0158)	(0.0590)	(0.0585)	(0.0591)	(0.0587)
State FE		✓	✓	✓	✓
Year FE		✓	✓	✓	✓
Race FE					
Cohort FE		✓			
Occupation FE				✓	✓
Industry FE			✓		✓
State FE Available	No	Yes	Yes	Yes	Yes
Year FE Available	No	Yes	Yes	Yes	Yes
Race FE Available	No	Yes	Yes	Yes	Yes
Cohort FE Available	No	Yes	Yes	Yes	Yes
Occupation FE Available	No	No	No	Yes	Yes
Industry FE Available	No	No	Yes	No	Yes
R-squared	0.013	0.019	0.028	0.034	0.036
N	32567	32556	25592	25426	25329

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$